

# Inferring microscopic kinetics of a Markov process using maximum caliber

Purushottam D. Dixit\*

*Department of Systems Biology, Columbia University, NY, NY 10032*

Ken A. Dill

*Laufer Center for Quantitative Biology,*

*Department of Chemistry,*

*and Department of Physics and Astronomy,*

*Stony Brook University, Stony Brook, NY, 11790*

We present a principled approach for estimating the matrix of microscopic rates among states of a Markov process, given only its stationary state population distribution and a single average global kinetic observable. We adapt Maximum Caliber, a variational principle in which a path entropy is maximized over the distribution of all the possible trajectories, subject to basic kinetic constraints and some average dynamical observables. We show that this approach leads, under appropriate conditions, to the continuous-time master equation and a Smoluchowski-like equation that is valid for both equilibrium and non-equilibrium stationary states. We illustrate the method by computing the solvation dynamics of water molecules from molecular dynamics trajectories.

---

\* pd2447@columbia.edu; Corresponding author

## I. INTRODUCTION

We are interested in a principled way to solve the following under-determined “inverse” kinetics problem. Consider a stationary and irreducible Markov process among  $i = 1, 2, 3, \dots, N$  states. Suppose you know: (a) the stationary state probability distribution,  $\{p_i\}$  of the occupancies of those states, and (b) the value  $\langle w \rangle$  of some dynamical observable  $w$  averaged over the ensemble of stationary state trajectories. From these  $N + 1$  quantities, we want to infer the  $N \times N$  *microscopic* transition rates,  $\{k_{ij}\}$  between those states.

This question is pertinent for situations such as the following. Often molecular dynamics simulations are performed on complex systems (1), where it the stable states can be sampled more efficiently than the transitions between them, because the latter involve crossing barriers that can sometime be high. Given the populations of the stable states, and a little experimental information about the overall rate of the process, it would be useful to estimate the full microscopic matrix of the transition rates between the states. Other examples include how amino-acid sequences of proteins change during the evolutionary dynamics of organisms, such as the HIV virus (2) and the collective firing patterns of neurons (3), etc.

Here, we propose a procedure based on the principle of Maximum Caliber, a variant of the principle of Maximum Entropy, that is applicable to dynamical processes (4-6).

First, we define *the path entropy*,  $\mathcal{S}$ , over a given ensemble  $\{\Gamma\}$  of trajectories  $\Gamma$  as:

$$\mathcal{S} = - \sum_{\{\Gamma\}} p(\Gamma) \log p(\Gamma). \quad (1)$$

Maximum Caliber is a variational principle that chooses a unique probability distribution  $\{P(\Gamma)\}$  over trajectories from all possible candidate distributions as the one that maximizes the path entropy while otherwise satisfying the relevant stationary and dynamical constraints (4, 6-8).

Consider an ensemble of stationary state trajectories  $\{\Gamma\}$  of a Markov process having a total time duration  $T$  where  $\Gamma \equiv \dots \rightarrow i \rightarrow j \rightarrow k \rightarrow l \rightarrow \dots$ . The Markov property implies that the probability of any particular trajectory  $\Gamma$  can be expressed in terms of the transition rates  $\{k_{ij}\}$ ,

$$P(\Gamma) = \dots k_{ij} \cdot k_{jk} \cdot k_{kl} \dots. \quad (2)$$

The path entropy of the above ensemble is directly proportional to the total duration  $T$  of the trajectory. The path entropy per unit time  $\mathcal{S}$  is given by (9)

$$\mathcal{S} = - \sum_{i,j} p_i k_{ij} \log k_{ij}. \quad (3)$$

The microscopic rates of any Markov process are subject to two types of constraints. First, from state  $i$  at time  $t$ , the system *must land somewhere* at time  $t + dt$ . Second, a system in state  $j$  at time  $t + dt$  *must arrive from somewhere*, so:

$$\sum_j k_{ij} = 1 \quad \forall i \text{ and } p_j = \sum_i p_i k_{ij} \quad \forall j \quad (4)$$

Third, we require one additional constraint that is *global*, i.e. averaged over the entire ensemble of trajectories. We fix the path ensemble average of some dynamical quantity  $w$ . The average  $\langle w \rangle_\Gamma$  over any given stationary trajectory  $\Gamma$  is given by

$$\langle w \rangle_\Gamma = \frac{1}{T} (\dots + w_{ij} + w_{jk} + w_{kl} + \dots). \quad (5)$$

The path ensemble average  $\langle w \rangle$  is

$$\langle w \rangle = \sum_\Gamma P(\Gamma) \langle w \rangle_\Gamma. \quad (6)$$

Since  $\Gamma$  is a stationary state trajectory, the path ensemble average  $\langle w \rangle$  of Eq. 6 simplifies to

$$\langle w \rangle = \sum_{i,j} p_i k_{ij} w_{ij}. \quad (7)$$

Maximization of the path entropy subject to these three constraints can be expressed equivalently in terms of maximization of a quantity called the Caliber  $\mathcal{C}$  (6):

$$\begin{aligned} \mathcal{C} = & - \sum_{i,j} p_i k_{ij} \log k_{ij} + \sum_i a_i \left( \sum_j p_i k_{ij} - p_i \right) \\ & + \sum_j l_j \left( \sum_i p_i k_{ij} - p_j \right) - \gamma \left( \sum_{i,j} p_i k_{ij} w_{ij} - \langle w \rangle \right) \end{aligned} \quad (8)$$

where  $\gamma$  is the Lagrange multiplier associated with the constraint  $\langle w \rangle$  and  $\{a_i\}$  and  $\{l_j\}$  enforce the *to-somewhere* constraint and the *from-somewhere* constraint, respectively.

To solve for the matrix  $k_{ij}$  of rates, we take the derivative of the Caliber  $\mathcal{C}$  with respect to  $k_{ij}$  and equate it to zero. This gives:

$$\begin{aligned} p_i (1 + \log k_{ij}^*) &= a_i p_i + l_j p_i - \gamma p_i w_{ij} \\ \Rightarrow k_{ij}^* &= \frac{\beta_i}{p_i} \lambda_j e^{-\gamma w_{ij}} \end{aligned} \quad (9)$$

where we have made the substitutions:  $e^{a_i-1} = \frac{\beta_i}{p_i}$  and  $e^{l_j} = \lambda_j$ . The values  $k_{ij}^*$  are the rates that satisfy the constraints and otherwise maximize the caliber. For simplicity of notation, we drop the superscript  $*$  in the remainder of this paper i.e.  $k_{ij}^* \equiv k_{ij}$ .

In this problem, the values of  $p_i$  are given. To compute the  $k_{ij}$ 's, we first must determine the values of the Lagrange multipliers  $\beta_i$ ,  $\lambda_j$ , and  $\gamma$ . We do so by substituting the constraint relations mentioned above.

### A. Determining the Lagrange multipliers

For a given value of  $\gamma$ , the *modified* Lagrange multipliers  $\beta_i$  and  $\lambda_j$  are determined by satisfying the *to-somewhere* and *from-somewhere* conditions indicated above. From Eqs. 4

$$\begin{aligned} 1 &= \sum_j \frac{\beta_i}{p_i} \lambda_j \mathcal{W}_{ij} \text{ and } p_j = \sum_i p_i \frac{\beta_i}{p_i} \lambda_j \mathcal{W}_{ij} \\ \Rightarrow \frac{p_i}{\beta_i} &= \sum_j \lambda_j \mathcal{W}_{ij} \text{ and } \frac{p_j}{\lambda_j} = \sum_i \beta_i \mathcal{W}_{ij} \end{aligned} \quad (10)$$

where  $e^{-\gamma w_{ij}} = \mathcal{W}_{ij}$ . Eq. 10 can be simplified if we define a non-linear operator  $\mathcal{D}$  over column vectors  $\bar{x} = [x_1, x_2, \dots]^T$  as  $\mathcal{D}[\bar{x}]_i = \frac{p_i}{x_i}$ . We have

$$\mathcal{W}\bar{\lambda} = \mathcal{D}[\bar{\beta}] \text{ and } \mathcal{W}^T\bar{\beta} = \mathcal{D}[\bar{\lambda}]. \quad (11)$$

where  $\bar{\lambda} = [\lambda_1, \lambda_2, \dots]^T$  and  $\bar{\beta} = [\beta_1, \beta_2, \dots]^T$  are the column vectors of Lagrange multipliers.

For a particular value of the Lagrange multiplier  $\gamma$ , Eqs. 11 can be numerically and self-consistently solved for  $\{\beta_i\}$  and  $\{\lambda_i\}$ . In practice, we choose an appropriate  $\gamma$  by first constructing transition rates  $\{k_{ij}\}$  for multiple values of  $\gamma$  (see Eq. 9) and choosing the value of  $\gamma$  which satisfies

$$\sum_{i,j} p_i k_{ij} w_{ij} = \sum_{i,j} \beta_i \lambda_j e^{-\gamma w_{ij}} w_{ij} = \langle w \rangle \quad (12)$$

where  $\langle w \rangle$  is the prescribed value of the ensemble average of the dynamical quantity  $w$ .

## II. AN ILLUSTRATION: COMPUTING THE DYNAMICS OF A SOLVATION SHELL FROM SIMULATED POPULATIONS.

We now illustrate how the present MaxCal method can be used to take a stationary-state distribution and a global dynamical constraint and to infer microscopic kinetics. Consider a shell of solvating water molecules surrounding a single water molecule. The number,  $n(t)$ , of water molecules in the hydration shell is a quantity that fluctuates with time  $t$  (see Fig. 1). We want to compute how fast the water molecules enter or exit the solvation shell. If the time interval  $dt$  is small,  $n(t)$  and  $n(t + dt)$  will be statistically correlated. Here, we construct a Markov process to model the time series  $\{n(t)\}$ . We will require the Markov process to reproduce a) the stationary distribution  $p(n)$  that is observed in molecular dynamics simulations, and b) the average change in occupancy  $\Delta$  per time step of duration  $dt$ , a path ensemble average. We have

$$\Delta = \langle |n(t + dt) - n(t)| \rangle = \sum_{i,j} |i - j| \cdot p(i) k_{ij} \quad (13)$$

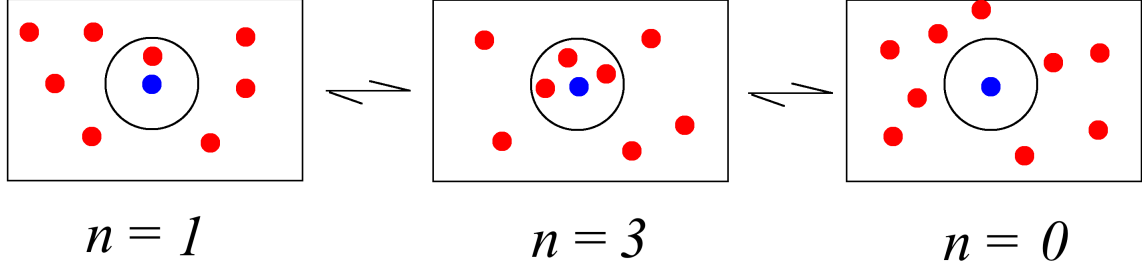


FIG. 1. The hydration shell (black circle) around a central water molecule (blue disc) is dynamically populated by other water molecules in the bulk solvent medium (red discs). The probability,  $p(n)$  that the hydration shell has exactly  $n$  water molecules is a key quantity in determining the solvation free energy of liquid water.

where  $n(t) = i$  and  $n(t + dt) = j$ .

From a trajectory sampled at every 5 fs from an MD simulation of liquid water (see appendix **III** for details of the simulation), we estimate the roughly normal stationary distribution  $p(n)$  (see in panel A of Fig. 2) (10, 11). When we constrain the observed stationary distribution  $p(n)$  and the mean jump size  $\Delta$ , the transition rate  $k_{ij}$  for a transition  $n(t) = i \rightarrow n(t + dt) = j$  is given by (see Eq. 9),

$$k_{ij} = \frac{\beta_i \lambda_j e^{-\gamma|i-j|}}{p(i)}. \quad (14)$$

For a given value of  $\gamma$ , we determine the Lagrange multipliers  $\beta_i$  and  $\lambda_i$  from Eqs. 11 above. In order to determine the Lagrange multiplier  $\gamma$  which dictates the rate of transition between states, we first construct Markov processes for different values of  $\gamma$ . Panel B of Fig. 2 shows that the path ensemble average of the change in occupation number per unit time step  $\Delta$  is exponentially decreasing with  $\gamma$ . From trajectories sampled at every 5 fs from the MD simulation, we find that experimental trajectory average  $\Delta_{\text{expt}} = \overline{|n(t + dt) - n(t)|} \approx 0.0629$  which corresponds to  $\gamma \approx 3.29$ . From here onwards, we use  $\gamma = 3.29$  and construct the transition rates  $\{k_{ij}\}$  (see Eq. 14). Note that the path ensemble average  $\Delta$  and consequently the Lagrange multiplier  $\gamma$ , depend on the time interval  $dt$  between two observation ( $dt = 5$  fs here).

From the Markov process constructed with  $\gamma = 3.29$  (see above), we now compute various dynamical quantities: a) the probability  $P_d$  of jump size  $d$ , b) the occupancy autocorrelation  $\langle \delta_{n(0), n(\tau)} \rangle$ , and c) the transition probabilities  $k_{ij}$ , and we compare to those obtained directly from the MD simulation trajectory. In general, the MaxCal method will be of value when rates are hard

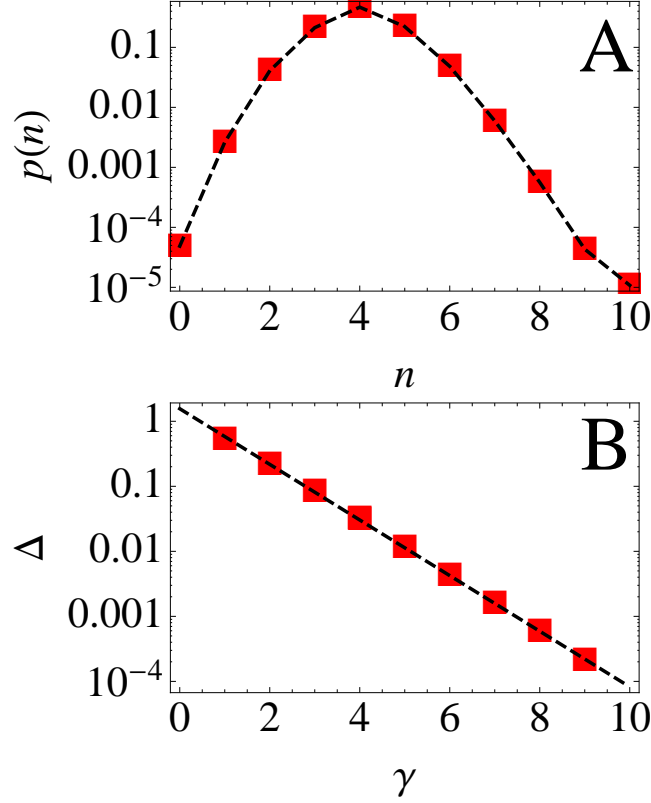


FIG. 2. Panel A: The stationary distribution  $p(n)$  of the number of water molecules in the hydration shell of radius  $r = 3.2$  Å of a reference water molecule. Panel B: The dependence of the ensemble average of change in water occupancy number  $\Delta = \langle |n(t + dt) - n(t)| \rangle$  on the Lagrange multiplier  $\gamma$ . We see that  $\Delta$  depends exponentially on  $\gamma$ . A higher  $\gamma$  implies slower dynamics and vice versa. We choose  $\gamma = 3.29$  to match the observed  $\Delta \approx 0.0629$  in the molecular dynamics simulation.

to simulate, such as for large kinetic barriers. Here, we are just illustrating with a toy problem for which we can determine the rates independently from the simulations.

From the long simulation trajectory, the probability  $P_d$  of jump size is estimated as the histogram of  $d = n(t + dt) - n(t)$ . Here  $d$  could be both positive and negative.  $P_d$  is given by

$$P_d = \sum_n p(n) k_{n,n+d}. \quad (15)$$

The normalized occupancy autocorrelation is simply the joint probability that  $n(t)$  and  $n(t + \tau)$  are equal. It is given by

$$\langle \delta_{n(0),n(\tau)} \rangle = \frac{\sum_n p(n) K(\tau)_{nn}}{\sum_n p(n)^2} \quad (16)$$

where  $K(\tau) = k^\tau$  is the  $\tau^{\text{th}}$  power of the matrix of transition rates  $\{k_{ij}\}$ .

In Fig. 3 we plot  $P_d$ ,  $\langle \delta_{n(0),n(\tau)} \rangle$ , and  $\{k_{ij}\}$  estimated from the molecular dynamics trajectory and compare them to our predictions from Markov modeling. Even though we constrained only

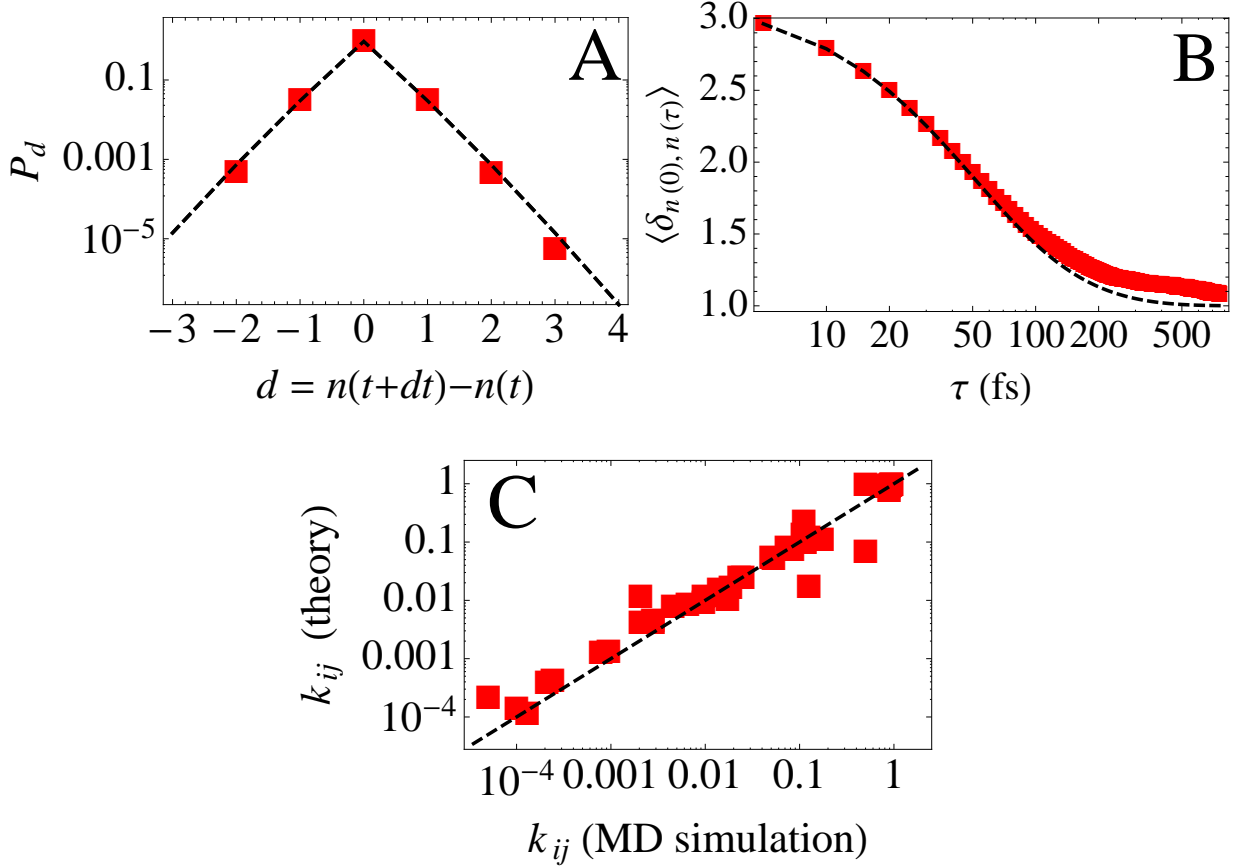


FIG. 3. Panel A: The probability  $P_d$  of jump size estimated from the trajectory derived from MD simulation (red squares) is compared to the one predicted using the transition rates of the Markov process (dashed black line). Panel B: The normalized occupancy autocorrelation  $\langle \delta_{n(0), n(\tau)} \rangle$  as estimated from the MD trajectory and as predicted from the transition rates of the Markov process. Panel C: We directly compare the transition rates  $k_{ij}$  for the probability of transition  $i \rightarrow j$  empirically obtained from the MD trajectory to the ones predicted by using Eq. 14.

the mean value  $\Delta = \langle |d| \rangle$  of  $P_d$ , the modeled Markov process captures the entire distribution  $P_d$  with high accuracy. Similarly the occupancy correlation  $\langle \delta_{n(0), n(\tau)} \rangle$  is also reproduced with high accuracy even though we did not utilize *any* information about it when inferring the transition rates of the Markov process. Moreover, our modeling also accurately captures the individual transition rates  $\{k_{ij}\}$  over 4 orders of magnitude.

### III. CONTINUOUS TIME AND CONTINUOUS SPACE LIMITS

Above, we have described discrete Markov processes. But, it is readily shown that the method can also be applied to continuous processes. Consider a Markov system whose states are points on a multi-dimensional discrete lattice with spacing  $dx$  in each dimension. Let  $d_{ij}$  be the manhattan distance between two states  $i$  and  $j$ . A wide variety of systems belong to this class including discretized brownian walks, lattice polymers, and spin glasses.

Assume that the system evolves continuously in time  $t$  but that we observe it only at a regular time interval  $dt \ll \tau$ , where  $\tau$  is the characteristic time constant of the system. Let us construct a Markov process by constraining the path ensemble average  $\langle d_{ij} \rangle$ , the average displacement per time step.

In appendix **I** we show that as  $dt \rightarrow 0$ , the Markov process described above is expressed by the master equation

$$\frac{d\bar{q}(t)}{dt} = -K^T \bar{q}(t). \quad (17)$$

The transition rates  $i \neq j$  are given by

$$K_{ij} = \mu \sqrt{\frac{p_j}{p_i}} \text{ if } d_{ij} = dx \text{ and } 0 \text{ if } d_{ij} > dx \quad (18)$$

$$K_{ii} = -\sum_{j \neq i} K_{ij}. \quad (19)$$

Here  $\bar{q}(t) = [q_1(t), q_2(t), \dots]$  denotes the instantaneous probability distribution. Intriguingly, the transition rates  $K_{ij}$  are functions of probability amplitudes  $\sqrt{p_j}$  and  $\sqrt{p_i}$ . We contrast this observation with the well known Glauber dynamics (12) for Ising-like systems where transtion rates are functions of probabilities themselves. Given that the continuous-space limit of the master equation Eq. 17 is the Smoluchowski equation (see below), we identify Eq. 17 as the unique form of the discrete-space Smoluchowski equation.

The continuous-time continuous space limit of Eq. 18 as  $dx \rightarrow 0$  is the so-called Smoluchowski equation for interacting degrees of freedom. In appendix **II** we show that Eq. 18 reduces to

$$\frac{\partial q(\bar{X}; t)}{\partial t} = D \nabla^2 q(\bar{X}; t) - D \nabla \cdot [q(\bar{X}; t) \cdot \bar{\mathcal{F}}(\bar{X})]. \quad (20)$$

Here,  $q(\bar{X}; t)$  is the instantaneous probability density of the state space point  $\bar{X}$ ,  $\phi(\bar{X}) = -\log p_{ss}(\bar{X})$  is the statistical field that corresponds to the stationary state density  $p_{ss}(\bar{X})$ , and  $D$  is an effective diffusion constant that sets the time scale for  $q(\bar{X}; t)$  to reach the steady state density  $p_{ss}(\bar{X})$ . We have denoted  $\bar{\mathcal{F}}(\bar{X}) = -\nabla \phi(\bar{X})$  as a restoring *force*. Note that Eq. 20 is valid even if  $\phi(\bar{X})$  is a dissipative field corresponding to a non-equilibrium steady state.



#### IV. DISCUSSION AND SUMMARY

We have presented here a variational approach that computes  $N \times N$  microscopic rate coefficients of a Markov process, given only knowledge of a stationary state population distribution and one trajectory-averaged dynamical property. In this approach, we maximize a path entropy subject to constraints. We show that this method correctly gives dynamical quantities on an example of molecular dynamics simulations of a water solvation shell around a single water molecule. This method may be useful for analyzing dynamical data from MD simulations (1), single-molecule experiments such as on ion channels (13), dynamics of neuron firing (3), and the dynamics of protein-sequence evolution (2), for example.

#### ACKNOWLEDGMENTS

KD would like to thank YYY. PD thanks Mr. Manas Rachh and Mr. Karthik Shekhar for numerous discussions about the topic and a critical reading of the manuscript.

- 
- [1] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, *et al.*, *Science* **330**, 341 (2010)
  - [2] K. Shekhar, C. F. Ruberman, A. L. Ferguson, J. P. Barton, M. Kardar, and A. K. Chakraborty, *Physical Review E* **0**, 0 (2013)
  - [3] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, *Nature* **440**, 1007 (2006)
  - [4] E. T. Jaynes, *Annual Review of Physical Chemistry* **31**, 579 (1980)
  - [5] E. Jaynes and D. Scalapino, unpublished preprint, available at: [bayes.wustl.edu/etj/articles/irreversible.stat.mac.pdf](http://bayes.wustl.edu/etj/articles/irreversible.stat.mac.pdf)(1963)
  - [6] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, *Rev. Mod. Phys.* **85**, 1115 (2013)
  - [7] G. Stock, K. Ghosh, and K. A. Dill, *J. Chem. Phys.* **128**, 194102 (2008)
  - [8] Q. A. Wang, *Chaos, Solitons & Fractals* **26**, 1045 (2005)
  - [9] T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley & Sons, 2012)
  - [10] D. Asthagiri, P. D. Dixit, S. Merchant, M. E. Paulaitis, L. R. Pratt, S. B. Rempe, and S. Varma, *Chem. Phys. Lett.* **485**, 1 (2010)
  - [11] S. Merchant and D. Asthagiri, *The Journal of chemical physics* **130**, 195102 (2009)
  - [12] R. J. Glauber, *Journal of mathematical physics* **4**, 294 (1963)
  - [13] B. Hille, *Ion channels of excitable membranes*, Vol. 507 (Sinauer Sunderland, MA, 2001)

- [14] W. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys **79**, 926 (1983)
- [15] E. Neria, S. Fischer, and M. Karplus, J. Chem. Phys **105**, 1902 (1996)
- [16] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, Journal of computational chemistry **26**, 1781 (2005)

## APPENDIX I: CONTINUOUS TIME LIMIT

Since Eq. 8 is convex in transition rates  $k_{ij}$ , there exists a unique matrix of transition rates  $\{k_{ij}\}$  that maximizes the Caliber among all transition rate matrices  $\{k_{ij}\}$  that satisfy the imposed stationary and dynamical constraints. In the main text, we showed that the transition rates are given by

$$k_{ij} = \frac{\beta_i}{p_i} \lambda_j e^{-\gamma w_{ij}} \quad (21)$$

where for a given value of  $\gamma$ , the Lagrange multipliers  $\{\beta_i\}$  and  $\{\lambda_j\}$  are determined by satisfying the normalization and stationary constraints of Eq. 11.

If the constrained quantity  $w_{ij}$  is symmetric in  $i$  and  $j$  for a transition  $i \rightarrow j$  i.e.  $w_{ij} = w_{ji} \forall i$  and  $j$ , we have

$$\mathcal{W}_{ij} = e^{-\gamma w_{ij}} = \mathcal{W}_{ji} = e^{-\gamma w_{ji}} \Rightarrow \mathcal{W} = \mathcal{W}^T \quad (22)$$

From Eq. 11 it follows that

$$\mathcal{W}\bar{\lambda} = \mathcal{D}[\bar{\beta}] \text{ and } \mathcal{W}\bar{\beta} = \mathcal{D}[\bar{\lambda}]. \quad (23)$$

Identifying  $\mathcal{W}^{-1}\mathcal{D} = \mathcal{G}$  as a non-linear operator on column vectors, we have

$$\bar{\lambda} = \mathcal{G}[\bar{\beta}] \text{ and } \bar{\beta} = \mathcal{G}[\bar{\lambda}]. \quad (24)$$

Notice that if  $\{\bar{\lambda}, \bar{\beta}\}$  is a solution of Eq. 24, so is  $\{\bar{\beta}, \bar{\lambda}\}$ . In other words, *since*  $\{k_{ij}\}$  are uniquely determined from the maximization of Eq. 8, we must have

$$\lambda_i \beta_j = \beta_i \lambda_j \forall i \text{ and } j. \quad (25)$$

From Eq. 25 it follows that  $\lambda_i = \eta \beta_i$  for some non-zero constant  $\eta$ . We have

$$\bar{\lambda} = \mathcal{W}^{-1}\mathcal{D}[\bar{\beta}] \Rightarrow \mathcal{G}[\bar{\beta}] = \eta \bar{\beta}. \quad (26)$$

In other words,  $\bar{\beta}$  is the eigenvector of the non-linear operator  $\mathcal{G} = \mathcal{W}^{-1}\mathcal{D}$  with eigenvalue  $\eta$  and  $\bar{\lambda} = \eta \bar{\beta}$ . Now it follows that

$$k_{ij} = \frac{\eta \lambda_i \lambda_j e^{-\gamma w_{ij}}}{p_i} \quad (27)$$

Now consider a discrete-time discrete-state Markov process over states  $\{i\}$ . Assume that the states are points on a multi-dimensional lattice with spacing  $dx$ . Let  $d_{ij}$  be the manhattan distance

between states  $i$  and  $j$ . Let us constrain the path ensemble average of the movement  $\langle d_{ij} \rangle$  in a single step over long trajectories. Since  $d_{ij}$  is symmetric, from Eq. 27 we have

$$k_{ij} = \frac{\eta \lambda_i \lambda_j e^{-\gamma d_{ij}}}{p_i}. \quad (28)$$

The Lagrange multipliers  $\bar{\lambda}$  are determined by solving Eq. 11

$$\mathcal{D}[\bar{\lambda}] = \eta \mathcal{W} \bar{\lambda}. \quad (29)$$

Applying  $\mathcal{D}$  on both sides and recognizing that  $\mathcal{D}[c\bar{x}] = \mathcal{D}[\bar{x}]/c$  for any non-zero  $c$ ,

$$\eta \bar{\lambda} = \mathcal{D}[\mathcal{W} \bar{\lambda}]. \quad (30)$$

Now consider the case where  $\gamma$  is positive and large we write  $\epsilon = e^{-\gamma dx} \ll 1$  where  $dx$  is the minimum pairwise separation between all states  $i$  and  $j$ . Here,  $\epsilon$  is small.

When  $\gamma$  is positive and large, the Markov process realistically only visits nearest neighbor states. For example, for any state  $i$ , the realistically accessible nearest neighbor states are all states  $j$  such that  $d_{ij} = dx$ .

Since  $d_{ii} = 0$ , we can write  $\mathcal{W} \approx \mathcal{I} + \epsilon \Delta$  where  $\mathcal{I}$  is the identity matrix.  $\Delta$  is a matrix of connectivity of nearest neighbors,  $\Delta_{ij} = 1$  iff  $d_{ij} = dx$  and zero otherwise. We have

$$\eta \bar{\lambda} = \mathcal{D}[\bar{\lambda} + \epsilon \bar{f}] \quad (31)$$

where

$$\bar{f} = \Delta \bar{\lambda}. \quad (32)$$

Expanding the right hand side of Eq. 31 and keeping terms of order up to  $\epsilon$ ,

$$\eta \lambda_i \approx \frac{p_i}{\lambda_i} - \frac{\epsilon p_i f_i}{\lambda_i^2}. \quad (33)$$

Notice that since  $\Delta_{ii} = 0$ ,  $f_i$  does not directly depend on  $\lambda_i$ . We solve Eq. 33 for  $\lambda_i$  and neglect terms of order higher than  $\epsilon$ . We have

$$\lambda_i \approx \sqrt{\frac{p_i}{\eta}} - \frac{\epsilon f_i}{2} \quad (34)$$

$$\Rightarrow \lambda_i = \frac{1}{\sqrt{\eta}} \left( \sqrt{p_i} - \frac{\epsilon}{2} \sum_j \Delta_{ij} \sqrt{p_j} \right) \quad (35)$$

Recall that

$$\begin{aligned} k_{ij} &= \frac{\eta \lambda_i \lambda_j e^{-\gamma d(i,j)}}{p_i} \\ &= \frac{\eta \epsilon \lambda_i \lambda_j}{p_i} \end{aligned} \quad (36)$$

$$\text{iff } d_{ij} = dx \text{ and } i \neq j \text{ and } 0 \text{ otherwise} \quad (37)$$

Substituting  $\lambda_i$  from Eq. 35 into Eq. 37 and retaining terms up to order 1 in  $\epsilon$ , we get

$$k_{ij} = \epsilon \sqrt{\frac{p_j}{p_i}} \quad \text{iff } d_{ij} = dx \text{ and } i \neq j \text{ and } 0 \text{ otherwise} \quad (38)$$

$$k_{ii} = 1 - \sum_{j \neq i} k_{ij} \quad (39)$$

Let  $\bar{q}(t)$  be the instantaneous probability distribution of the Markov process. From Eq. 39, we have

$$q_j(t + dt) = \sum_i k_{ij} q_i(t). \quad (40)$$

Writing  $\epsilon = \mu dt$  and taking the limit  $dt \rightarrow 0$ , we derive Eq. 17.

## APPENDIX II: DERIVING THE SMOLUCHOWSKI EQUATION

Here, we will derive the Smoluchowski equation for a particle diffusing in a one dimensional landscape. The generalization Eq. 20 presented in the main text is trivially obtained from the one-dimensional equation. Let us consider a particle moving between  $x = -L$  and  $x = L$  in discrete steps of size  $dx$ . Let the stationary distribution  $p_{ss}(x)$  be governed by a potential  $\phi(x)$  such that  $p_{ss}(x) = e^{-\phi(x)}$ . Let us assume that the particle diffuses in continuum time on this discrete landscape. Denote the instantaneous probability density by  $q(x; t)$ . From Eq. 18, we have

$$\begin{aligned} \frac{dq(x; t)}{dt} = & \underbrace{\mu \left( q(x + dx; t) \sqrt{\frac{e^{-\phi(x)}}{e^{-\phi(x+dx)}}} + q(x - dx; t) \sqrt{\frac{e^{-\phi(x)}}{e^{-\phi(x-dx)}}} \right)}_{\text{entry into state } x \text{ from its neighbors}} \\ & - \underbrace{\mu \left( q(x; t) \sqrt{\frac{e^{-\phi(x+dx)}}{e^{-\phi(x)}}} + q(x; t) \sqrt{\frac{e^{-\phi(x-dx)}}{e^{-\phi(x)}}} \right)}_{\text{leakage from state } x \text{ to its neighbors}} \end{aligned} \quad (41)$$

The first term in Eq. 41 corresponds to the probability flow into state  $x$  from its neighbors  $x - dx$  and  $x + dx$  while the second terms corresponds to flow out of state  $x$  into its neighbors  $x - dx$  and  $x + dx$ .

In order to see the continuous-space limit of Eq. 41, assume that  $q(x; t)$  and  $\phi(x)$  are differentiable functions in  $x$ . Expanding the right hand side of Eq. 41 as a Taylor series up to two orders

in  $dx$ , we have

$$\begin{aligned}\frac{\partial q(x;t)}{\partial t} &= \mu dx^2 \left( \frac{\partial^2 q(x;t)}{\partial x^2} + \frac{\partial}{\partial x} \left[ q(x;t) \cdot \frac{d\phi(x)}{dx} \right] \right) \\ &= D \left( \frac{\partial^2 q(x;t)}{\partial x^2} - \frac{\partial}{\partial x} [q(x;t) \cdot \mathcal{F}(x)] \right)\end{aligned}\tag{42}$$

where  $D = \mu dx^2$  is the diffusion constant and we have identified  $\mathcal{F}(x) = -\frac{d\phi(x)}{dx}$  as a *restoring force*. The continuum-limit exists only when the rate  $\mu$  scales such that  $\mu dx^2$  is constant. Eq. 42 is valid not only for equilibrium situations ( $\phi(x) = \beta G(x; \beta)$ , where  $G(x; \beta)$  is a free energy landscape and  $\beta$  is inverse temperature) but also for non-equilibrium steady states (NESS).

### APPENDIX III: MD SIMULATION

We performed a molecular dynamics simulation on 233 water molecules (14, 15) at 300K and at a constant volume using NAMD (16) with help of the Langevin thermostat. The oxygen atom of one of the water molecules was fixed at the origin. The time step of integration was 1 fs and the trajectory was stored every 5 fs. Sampling the trajectory every 5 fs ensures that correlations in  $n(t)$  haven't vanished.